

ABSTRACT OF THE DISCLOSURE

A system and method for automatic and semi-automatic document indexing preferably performs document recognition procedures on scanned document data and searches for sub-section delimiters. An index or table of contents is then generated for the document based on sub-section delimiters located in the search. For example, text strings, font size, or other symbols or distinguishing characteristics are used as delimiters in order to automatically find chapter or sub-section headings. A system operator may make adjustments to the automatically determined subdivisions. Alternatively, a plurality of document pages is displayed, for example, in thumbnail form, and the system operator indicates subdivision demarcation points within the displayed thumbnails. The indicated demarcation points are used as sub-section delimiters, and document subdivision is performed as described above. In yet another alternative, demarcation symbols are added to the document prior to scanning.